What is claimed is:

1. A method for indexing documents in a collection of documents, each document comprising one or more index terms, the method comprising:

determining a value x such that at least a majority of the index terms occur in x documents or fewer;

determining a value y, where y is not equal to x;

generating an inverted index for the collection of documents, the inverted index including an inverted list for each of the index terms, each inverted list including at least one posting and, if the number of postings exceeds x, further including a skip entry after the x^{th} posting and one or more skip entries thereafter at intervals of every y^{th} posting.

2. The method of claim 1, wherein each posting includes a document identifier identifying a document in the collection of documents, a position identifier identifying a position of the index term in the document, and a frequency of the index term occurring in the document.

3. The method of claim 1, wherein a skip entry identifies the smallest document number of documents included in the postings immediately following the skip entry in the inverted list.

4. The method of claim 3, wherein the skip entry further includes information to locate the next skip entry in the inverted list.

5. The method of claim 1, wherein a skip entry identifies the largest document number of documents included in the postings immediately preceding the skip entry in the inverted list.

6. The method claim 5, wherein the skip entry further includes information to locate the next skip entry in the inverted list.

7. The method of claim 1, wherein y is less than x.

8. The method of claim 1, wherein x is in the range of 256 to 512 and y is in the range of 128 to 256.

9. The method of claim 1, wherein the collection of one or more documents includes one or more binary files, data tables, source code files, text documents or combinations thereof.

10. The method of claim 1, further comprising:

compressing the inverted index.

11. The method of claim 1, wherein substantially all of the index terms occur in x documents or fewer.

12. The method of claim 11, wherein approximately 80 to 90% of the index terms occur in x documents or fewer.

13. The method of claim 1, wherein for each inverted list, if the number of postings exceeds x, further including a skip entry before the first posting in the inverted list.

14. The method of claim 1, wherein for each inverted list, if the number of postings exceeds x, further including a skip entry after the last posting in the inverted list.

15. A method for indexing documents, the method comprising:

receiving a collection of documents, each document comprising one or more index terms;

determining a value x, wherein at least a majority of the index terms occur in x documents or fewer and x is in the range of 256 to 512;

determining a value y, wherein y is not equal to the value x and is in the range of 128 to 256; and

generating an inverted index for the collection of documents, the inverted index including an inverted list for each of the index terms, each inverted list including at least one posting and, if the number of postings exceeds x, further including a skip entry after the x^{th} posting and one or more skip entries thereafter at intervals of every y^{th} posting.

16. The method of claim 15, wherein each posting includes a document identifier identifying a document in the collection of documents, a position identifier identifying a position of the index term in the document, and a frequency of the index term occurring in the document.

17. The method of claim 15, wherein a skip entry identifies the smallest document number of documents included in the postings immediately following the skip entry in the inverted list.

18. The method of claim 17, wherein the skip entry further includes information to locate the next skip entry in the inverted list.

19. The method of claim 15, wherein a skip entry identifies the largest document number of documents included in the postings immediately preceding the skip entry in the inverted list.

20. The method claim 15, wherein the skip entry further includes information to locate the next skip entry in the inverted list.

21. The method of claim 15, wherein substantially all of the index terms occur in x documents or fewer.

22. The method of claim 21, wherein approximately 80 to 90% of the index terms occur in x documents or fewer.

23. The method of claim 15, wherein for each inverted list, if the number of postings exceeds x, further including a skip entry before the first posting in the inverted list.

24. The method of claim 15, wherein for each inverted list, if the number of postings exceeds x, further including a skip entry after the last posting in the inverted list.

25. An inverted index for a collection of documents, each document comprising one or more index terms, the inverted index comprising:

an inverted list for each index term in the collection of documents; and

one or more inverted lists including a quantity of postings that exceeds a value x, a skip entry after the x^{th} posting, and one or more additional skip entries thereafter at intervals of every y^{th} posting, where the value x is such that at least a majority of the index terms occur in x documents or fewer, and the value y is not equal to the value x.

26. The inverted index of claim 25, wherein each posting includes a document identifier identifying a document in the collection of documents, a position identifier identifying a position of the index term in the document, and a frequency of the index term occurring in the document.

27. The inverted index of claim 25, wherein a skip entry identifies the smallest document number of documents included in the postings immediately following the skip entry in the inverted list.

28. The inverted index of claim 27, wherein the skip entry further includes information to locate the next skip entry in the inverted list.

29. The inverted index of claim 25, wherein a skip entry identifies the largest document number of documents included in the postings immediately preceding the skip entry in the inverted list.

30. The inverted index of claim 29, wherein the skip entry further includes information to locate the next skip entry in the inverted list.

31. The inverted index of claim 25, wherein x is in the range of 256 to 512 and y is in the range of 128 to 256.

32. The inverted index of claim 25, wherein substantially all of the index terms occur in x documents or fewer.

33. The inverted index of claim 32, wherein approximately 80 to 90% of the index terms occur in x documents or fewer.

34. The inverted index of claim 25, wherein the collection of one or more documents includes one or more binary files, data tables, source code files, text documents or combinations thereof.

35. The inverted index of claim 25, wherein the one or more inverted lists further include a skip entry before the first posting in the inverted list.

36. The inverted index of claim 25, wherein the one or more inverted lists further include a skip entry after the last posting in the inverted list.

37. An article comprising a machine-readable medium storing instructions operable to cause one or more machines to perform operations comprising:

determining a value x such that at least a majority of the index terms occur in x

documents or fewer;

determining a value y, where y is not equal to x;

generating an inverted index for the collection of documents, the inverted index including an inverted list for each of the index terms, each inverted list including at least one posting and, if the number of postings exceeds x, further including a skip entry after the x^{th} posting and one or more skip entries thereafter at intervals of every y^{th} posting.

38. The article of claim 37, wherein each posting includes a document identifier identifying a document in the collection of documents, a position identifier identifying a position of the index term in the document, and a frequency of the index term occurring in the document.

39. The article of claim 37, wherein a skip entry identifies the smallest document number of documents included in the postings immediately following the skip entry in the inverted list.

40. The article of claim 39, wherein the skip entry further includes information to locate the next skip entry in the inverted list.

41. The article of claim 37, wherein a skip entry identifies the largest document number of documents included in the postings immediately preceding the skip entry in the inverted list.

42. The article of claim 41, wherein the skip entry further includes information to locate the next skip entry in the inverted list.

43. The article of claim 37, wherein y is less than x.

44. The article of claim 37, wherein x is in the range of 256 to 512 and y is in the range of 128 to 256.

45. The article of claim 37, wherein the collection of one or more documents includes one or more binary files, data tables, source code files, text documents or combinations thereof.

46. The article of claim 37, further comprising instructions operable to cause one or more machines to perform operations comprising:

compressing the inverted index.

47. The article of claim 37, wherein substantially all of the index terms occur in x documents or fewer.

48. The article of claim 47, wherein approximately 80 to 90% of the index terms occur in x documents or fewer.

49. The article of claim 37, wherein for each inverted list, if the number of postings exceeds x, further including a skip entry before the first posting in the inverted list.

50. The article of claim 37, wherein for each inverted list, if the number of postings exceeds x, further including a skip entry after the last posting in the inverted list.

51. An article comprising a machine-readable medium storing instructions operable to cause one or more machines to perform operations comprising:

receiving a collection of documents, each document comprising one or more index terms;

determining a value x, wherein at least a majority of the index terms occur in x documents or fewer and x is in the range of 256 to 512;

determining a value y, wherein y not equal to the value of x and is in the range of 128 to 256;

generating an inverted index for the collection of documents, the inverted index including an inverted list for each of the index terms, each inverted list including at least one posting and, if the number of postings exceeds x, further including a skip entry after the x^{th} posting and one or more skip entries thereafter at intervals of every y^{th} posting.

52. The article of claim 51, wherein each posting includes a document identifier identifying a document in the collection of documents, a position identifier identifying a position of the index term in the document, and a frequency of the index term occurring in the document.

53. The article of claim 51, wherein a skip entry identifies the smallest document number of documents included in the postings immediately following the skip entry in the inverted list.

54. The article of claim 53, wherein the skip entry further includes information to locate the next skip entry in the inverted list.

55. The article of claim 51, wherein a skip entry identifies the largest document number of documents included in the postings immediately preceding the skip entry in the inverted list.

56. The article of claim 51, wherein the skip entry further includes information to locate the next skip entry in the inverted list.

57. The article of claim 51, wherein substantially all of the index terms occur in x documents or fewer.

58. The article of claim 57, wherein approximately 80 to 90% of the index terms occur in x documents or fewer.

59. The article of claim 51, wherein for each inverted list, if the number of postings exceeds x, further including a skip entry before the first posting in the inverted list.

60. The article of claim 51, wherein for each inverted list, if the number of postings exceeds x, further including a skip entry after the last posting in the inverted list.